FE 515 Midterm

Oct 2018

This is an open-book midterm, you can use all materials that are available to you. However, any kind of communication tools is not allowed. Please notice that this midterm can be solved in multiple approaches. Regardless of the method you would use, you are required to answer all questions.

For this midterm, **10 points** is reserved for the answers' formatting. For this part you are required to use the Google's R style as discussed in class. Your submission should be easy to understand and clear to read.

Question 1 (20 Points)

In the board game monopoly, a player tosses a fair 6 sides dice and the point shown on the dice is the number of steps he or she should go. For example a player gets a 2, then he or she should go 2 steps forward.

1. What is the probability to get the first "6" at the sixth toss? Use dgeom() to answer this question.
2. If you toss this dice three times in a row. What is the probability to see two consecutive same number? Repeat this action 1000 time and calculate the probability based on your observation.

Question 2 (70 Points)

XLF is the ticker of Financial Select Sector SPDR ETF. Now, I want to use linear regression model to replicate the movement of XLF. Here is the list of possible equities:

- AAPL, BAC, C, DIS, F, GS, JPM, MS, PG, TWTR, WFC

To answer this question, you need to:

1. Download daily data for all the tickers (12 in total) from Yahoo finance. Then calculate daily return (either type) for each equity. In this question, you need to download 3-year-length of data.
2. Based on the returns you calculated, which equities are normal distributed? Report your finding in the R code.
3. Split your data set from Step 1 into two parts. The first part should contain 2-year-length of data (training data), the second part has the last year of data (testing data).
4. Build a multiple linear regression model based on the training data. In this model, all factors should be significant. In the end, write down the expression of linear regression model in the R code.
5. Based on your model, which equity is the most important one? Rank the ticker name from high impact to low impact.
6. Generate a plot for XLF. In this plot, you should use the time as X-axis input. Meanwhile, you need to show following information:
    - Plot the daily low price line and daily high price line.

- Fill the spread between low price and high price using gray color.
- Add the daily close price line (not adjusted close) to the plot.
- Add a legend in your plot.

Your plot should be easy and clear to understand.

Bonus 1: 20 Points

Question 2 can be rewrite into a function. In this bonus question, please create a self-defined function which is capable to handle Step 1, 3, 4, 5 in Question 2.

Function inputs:

- Ticker list
- Start time
- End time

Function output:

- Ranking of the ticker.

Grading details:

1. Your function can handle Step 1 (4 points)
2. Additionally, your function can handle 3 (4 points)
3. Additionally, your function can handle 4 (8 points)
4. Additionally, your function can handle 5 (4 points)

Bonus 2: 30 Points

This bonus question is the extension of Question 2. In order to answer this question, you may need to finish Question 2 step 1 to step 4. (This bonus questions contains content we haven't taught yet)

In step 4, you already have the linear model. Now, we want to test whether our model is useful.

1. Use the model from Step 4 and testing data from Step 3 to generate an estimated XLF return sequence. (10 points)
2. T-test is one method to compare the mean value from two groups. When the mean value of two groups are indifferent, we think both groups are similar. Perform T-test to compare the estimated return and XLF historical return. Do you think the estimated results are indifferent from historical data? (20 points)